数理解析研究所講究録
507巻 1983年 1-15

A CHARACTERIZATION OF SECOND ORDER EFFICIENCY FOR ESTIMATORS IN A CURVED EXPONENTIAL FAMILY

By Shinto Eguchi (江口 真透)

Osaka University

Asymptotic properties of estimators are considered in an m-dimensional curved exponential family $\tilde{\mathcal{F}}$ which is embedded in an exponential family of dimension n. It is shown that any first order efficient estimator is induced to a unique right triangle with sides $\sqrt{m/2}$, $\sqrt{(n-m)/2}$ and $\sqrt{n/2}$. Let L(u) be the likelihood function of a sample of size N with respect to an m-component parameter u describing $\tilde{\mathcal{F}}$. A necessary and sufficient condition for second order efficiency of an estimator $\hat{u}$ is given by

$$\lim_{N\to\infty} E\left[\frac{L(\hat{u})}{L(\tilde{u})}\right]^N \geqq 1$$

for any first order efficient estimator $\tilde{u}$. The condition implies second order efficiency of the maximun likelihood estimator which is famous as Fisher-Rao's theorem.

AMS subject 80 60F, 62F.
Key words and phrases. almost-metric structure, contrast function, curved exponential family, exponential family, Kullback-Leibler divergence, maximum likelihood estimator, second order efficiency, second fundamental tensor.

1. Introduction and main results. Let $\mathfrak{F}$ be an n-dimensional exponential family of densities on the data-space $\mathbb{R}^n$ with respect to a carrier measure ω. The family $\mathfrak{F}$ is expressed as

$$\{f(x \mid \theta) \equiv e^{<x,\theta>-\psi(\theta)} \quad : \theta \in \text{Ⓗ} \}$$

by the natural co-ordinate system $\theta \equiv (\theta^1,\cdots,\theta^n)$ with the usual inner product $<\cdot\ ,\ \cdot>$ of $\mathbb{R}^n$. The dual co-ordinates $\eta \equiv (\eta_1,\eta_2,\cdots,\eta_n)$ of $\mathfrak{F}$ is defined by the transformation of θ into η:

$$\eta[\theta] \equiv E_\theta\ x.$$

Then the maximum likelihood estimator of η or θ based on a sample $(x_1,x_2,\cdot\cdot,x_N)$ is given by

$$\hat{x} \equiv \frac{1}{N}(x_1 + x_2 + \cdots + x_N)$$

or $\hat{\theta} \equiv \theta[\hat{x}]$, respectively, where $\theta[\cdot]$ denotes the inverse transformation of $\eta[\cdot]$.

An m-dimensional curved exponential family is denoted by $\tilde{\mathfrak{F}}$ $(m < n)$, i.e.,

$$\tilde{\mathfrak{F}} \equiv \{f(x \mid \theta(u)) : u \in U\} ,$$

where U is an open set in $\mathbb{R}^m$ and the map $\theta(\cdot)$ from U to Ⓗ is nonlinear with the Jacobian matrix of rank m on U. Let $(x_1,x_2,\cdots, x_n)$ be an i.i.d. sample from a density $f(\cdot|\theta(u))$. We may confine estimators of u to the form of mappings of $\hat{x}$ or $\hat{\theta}$ since each of statistics $\hat{x}$ and $\hat{\theta}$ is minimal sufficient owing to the nonlinearity of $\theta(\cdot)$. Fisher-consistency of an estimator $\hat{u} = \hat{u}(\hat{\theta})$ is defined by

$$\hat{u}(\theta(u)) = u$$

for all u in U. For an estimator $\hat{u}$, $\Delta_N(\hat{u}, u)$ denotes the difference between the information matrix of the sample and that of the estimator, which is called the information loss incured by $\hat{u}$. A Fisher-consistent estimator $\hat{u}$ is said to be first order efficient if

$$\lim_{N\to\infty} \frac{1}{N} \Delta_N (\hat{u}, u) = 0.$$

Furthermore a first order efficient estimator $\hat{u}$ is said to be second order efficient if

$$\lim_{N\to\infty} [\Delta_N (\tilde{u}, u) - \Delta_N (\hat{u}, u)] \geqq 0$$

for all first order efficient estimator $\tilde{u}$, where $M \geqq 0$ denotes the non-negative definiteness of M. The Kullback-Leibler divergence $\rho_{KL}(f_1, f_2)$ between f_1 and f_2 in $\mathcal{F}$ is expressed as

$$\rho_{KL} (\theta_1, \theta_2) \equiv < \eta[\theta_1], \theta_1-\theta_2 > - \psi(\theta_1) + \psi(\theta_2)$$

with respect to θ, where $f_p = f(\cdot|\theta_p)$ with p= 1,2.
The following theorems 1, 2 and 3 will be proved in Section 2.

THEOREM 1. First order efficiency of a Fisher-consistent estimator $\hat{u}$ is equivalent to each of the following conditions (i), (ii) and (iii);

(i) $\lim_{N\to\infty} N \, E \, [\rho_{KL}(\hat{\theta},\theta(\tilde{u})) - \rho_{KL}(\hat{\theta}, (\hat{u}))] \geqq 0$

for any Fisher-consistent estimator $\tilde{u}$.

(ii) $\lim_{N\to\infty} N \, E \; \rho_{KL}(\theta(\hat{u}), \theta(u)) = m/2.$

(iii) $\lim\limits_{N\to\infty} N\,E\,\rho_{KL}(\hat{\theta},\theta(u)) = (n-m)/2.$

THEOREM 2 enables us to associate the common property of all first order efficient estimators with a right triangle, since the Kullback-Leibler divergence is the same order as the squared distance of $\mathfrak{F}$ (see Figure).

<<<<<Figure>>>>>

The measure

$$N\,E\,[\rho_{KL}(\hat{\theta},\theta(\hat{u})) - \rho_{KL}(\hat{\theta},\theta(\tilde{u}))]$$

is closely related to the discrimination rate of ρ_{KL}, introduced by Kuboki [5], in the model $\tilde{\mathfrak{F}}$, i.e., the case including the sufficient statistic $\hat{\theta}$. However we, here, consider this as a criterion between estimators $\hat{u}$ and $\tilde{u}$.

Let $L(u)$ be the likelihood function based on the sample $(x_1,\cdots,x_N)$. Since we have the relation

(1.1) $\log L(u_1) - \log L(u_2) = N\{\rho_{KL}(\hat{\theta},\theta(u_2)) - \rho_{KL}(\hat{\theta},\theta(u_1))\}$

for all u_1 and u_2 in U, THEOREM 1 can be rewritten as

COROLLARY 1. A Fisher-consistent estimator $\hat{u}$ is first order efficient if and only if

$$\lim_{N\to\infty} E\left[\frac{L(\hat{u})}{L(\tilde{u})}\right] \geqq 1$$

for all Fisher-consistent estimators $\tilde{u}$.

Moreover we shall show

THEOREM 2. A first-order efficient estimator $\hat{u}$ is second order efficient if and only if

$$(1.2) \quad \lim_{N\to\infty} E\, N^2 \, [\rho_{KL}(\hat{\theta},\theta(\tilde{u})) - \rho_{KL}(\hat{\theta},\theta(\hat{u}))] \geqq 0$$

for all first order efficient estimators $\tilde{u}$.

THEOREM 2 does not hold if the relation (1.2) is replaced by

$$\lim_{N\to\infty} E\, N^2 \, [\rho_{KL}(\theta(\tilde{u}),\theta(u)) - \rho_{KL}(\theta(\hat{u}),\theta(u))] \geqq 0.$$

This phenomenon comes from the naming term by the parametrization, which may be similar to the discussion of the mean squared errors for estimators (c.f. Rao [8], Efron [3] and Amari [1]).

The relation (1.1) leads us directly to

COROLLARY 2. Second order efficiency of a first order efficient $\hat{u}$ is equivalent to the condition:

$$\lim_{N\to\infty} E \left[\frac{L(\hat{u})}{L(\tilde{u})} \right]^N \geqq 1$$

for all first order efficient estimators $\tilde{u}$.

By definition, the maximum likelihood estimator $\hat{u}_{ML}$ satisfies

$$\frac{L(\hat{u}_{ML})}{L(\tilde{u})} \geqq 1$$

for any $\tilde{u}$ in U and any sample size N. So COROLLARY 2 implies promptly in second order efficiency of the maximum like likelihood estimator, which is famous for Fisher-Rao's theorem.

A contrast function ρ on $\mathfrak{F}$ is defined by satisfying the following conditions for any f_1 and f_2 in $\mathfrak{F}$:

(i) $\rho(f_1, f_2) \geqq 0$

(ii) $\rho(f_1, f_2) = 0 \Longleftrightarrow f_1 = f_2$ a.e. ω.

Dawid-Amari's almost-metric structure is denoted by A, and the almost-metric structure associated with ρ is denoted by $A(\rho)$ (see Appendix I).

There may remain a question of whether this criterion

$$E\ \rho_{KL}(\hat{\theta},\ \theta(\cdot))$$

is favorable only to the maximum likelihood estimator $\hat{u}_{ML}$ since the minimum Kullback-Leibler divergence estimator is nothing but $\hat{u}_{ML}$. However this question will vanish by

THEOREM 3. Let ρ be a contrast function with the almost-metric structure $A(\rho)$. The condition (1.2) for ρ in place of ρ_{KL} holds if $A(\rho) = A$ on $\tilde{\mathfrak{F}}$.

2. Proofs of the results.

We adopt the differential geometric formulation, due to Amari [1], including the almost-metric structure $A = (g, \overset{m}{\Gamma}, \overset{e}{\Gamma})$ over $\mathfrak{F}$.

For any f in $\tilde{\mathfrak{F}}$, let $T_f(\mathfrak{F})$ be the tangent space of $\tilde{\mathfrak{F}}$ at f. $T_f(\mathfrak{F})$ is decomposed into the tangent and normal spaces of $\tilde{\mathfrak{F}}$ at f with respect to the information metric g, i.e.,

$$T_f(\mathfrak{F}) = T_f(\tilde{\mathfrak{F}}) + T_f^{\perp}(\tilde{\mathfrak{F}}).$$

An $n\times(n-m)$ matrix $[B_\lambda^i(u)]_{\substack{i=1,2,\cdots,n\\ \lambda=m+1,\cdots,n}}$ can be chosen to satisfy

(2.1) $\quad B_\lambda^i(u)\ g_{ij}(\theta(u))\ B_a^j(u) = 0$

for $a = 1,2,\ldots,m$, where $B_a^i(u) = \partial\theta^i(u)/\partial u^a$. In the sequal we use the summation convention as in (2.1). With respect to co-ordinates θ and u, the bases of $T_f(\mathfrak{F})$, $T_f(\tilde{\mathfrak{F}})$ and $T_f(\tilde{\mathfrak{F}})$ at $f = f(\cdot|\theta(u))$ are represented as

$$\{e_i(u) \equiv \hat{x}_i - \eta_i(u)\}_{i=1,2,\ldots,n},$$
$$\{e_a(u) \equiv B_a^i(u)\ e_i(u)\}_{a=1,2,\ldots,m}$$

and

$$\{e_\lambda(u) \equiv B_\lambda^i(u)\ e_i(u)\}_{\lambda=m+1,\ldots,n},$$

respectively, where $\eta(u) \equiv \eta[\theta(u)]$. The induced components of g to $T_f(\tilde{\mathfrak{F}})$ and $T_f(\tilde{\mathfrak{F}})$ at $f = f(\cdot|\theta(u))$ are expressed as

$$\tilde{g}_{ab}(u) \equiv B_a^i(u)\ g_{ij}(\theta(u))\ B_b^j(u)$$

and

$$\tilde{g}_{\lambda\mu}(u) \equiv B_\lambda^i(u)\ g_{ij}(\theta(u))\ B_\mu^j(u),$$

respectively.

The second fundamental tensors of $\tilde{\mathfrak{F}}$ with respect to $\overset{m}{\Gamma}$ and $\overset{e}{\Gamma}$ are denoted by $\overset{m}{H}$ and $\overset{e}{H}$, respectively. The components of $\overset{m}{H}$ and $\overset{e}{H}$ are expressed as

$$\overset{m}{H}_{ab\lambda}(u) = B_\lambda^i(u)\ \partial_a[B_b^j(u)\ g_{ji}(\theta(u))]$$

and

$$\overset{e}{H}_{ab\lambda}(u) = B_\lambda^i(u)\ g_{ij}(\theta(u))\ \partial_a\ B_b^j(u)$$

with respect to u with $\partial_a \equiv \partial/\partial u^a$. Henceforth we omit the arguments of the above geometric quantities at the true value u and freely raise or lower indices of them, e.g.,

$$\overset{m}{H}{}_{b\lambda}^{a} \equiv \overset{m}{H}_{cb\lambda}(u)\ \tilde{g}^{ca}(u)$$

and

$$e^\lambda \equiv \tilde{g}^{\lambda\mu}(u)\ e_\mu(u),$$

where $\tilde{g}^{ca}(u)$ and $\tilde{g}^{\lambda\mu}(u)$ are the inverse elements of $\{\tilde{g}_{ac}(u)\}_{a,c=1,2,\ldots,m}$ and $\{\tilde{g}_{\mu\lambda}(u)\}_{\lambda,\mu=m+1,\ldots,n}$ respectively. For a first order efficient estimator $\hat{u}(\hat{\theta})$, the set

$$\{f(\cdot|\theta);\ \hat{u}(\theta) = u\}$$

is called the ancillary subspace of $\hat{u}$ of which the second fundamental tensor at $f = f(\cdot|\theta(u))$ with respect to $\overset{m}{\Gamma}$ is denoted by $\hat{H}$. Then we can rewrite THEOREM 7 in Amari [1] in the following convenient form:

THEOREM A. Let $\hat{u}$ be a first order efficient estimator of u. Then

$$\hat{u}^a - u^a = e^a - \frac{1}{2}\overset{m}{\Gamma}{}^a_{bc}e^b e^c + \overset{e}{H}{}^a_{b\lambda}e^b e^\lambda - \frac{1}{2}\hat{H}{}^a_{\lambda\mu}e^\lambda e^\mu + O(||e||^3). \tag{2.2}$$

Furthermore the estimator $\hat{u}$ is second order efficient if and only if the tensor $\hat{H}$ vanishes on $\tilde{\mathfrak{g}}$.

In practice THEOREM A is also equivalent to THEOREM 1 (ii) by Ghosh and Subramanyan [7] in the case of one parameter. We set about, on the basis of THEOREM A and Appendix II,

PROOF of THEOREM 1. The Kullback-Leibler divergence ρ_{KL} can be also expressed as

$$\rho_{KL}(\eta_1,\eta_2) = \langle\eta_1,\theta[\eta_2]-\theta[\eta_2]\rangle - \psi(\theta[\eta_1]) + \psi(\theta[\eta_2])$$

with respect to η. Let $\hat{u}$ be a first order efficient estimator. Then the statistic $\rho_{KL}(\hat{x}, \eta(\hat{u}))$ can be expanded as

$$\rho_{KL}(\hat{x}, \eta(\hat{u})) = \frac{1}{2}\, e_i(\hat{u})e_j(\hat{u})g^{ij} + O(||e||^3), \tag{2.3}$$

where g^{ij} is the inverse element of $\{g_{ij}\}$. It follows from THEOREM A that

$$e_i(\hat{u}) = B_{i\lambda}e^\lambda + O(||e||^2).$$

Hence we have from Appendix II that

$$\lim_{N\to\infty} N\ E\ \rho_{KL}(\hat{x}, \eta(\hat{u})) = g_{\lambda\mu}g^{\lambda\mu} = (n-m)/2$$

The rest of the assertions are similar to the above and so we complete the proof.

To prove THEOREM 2, we prepare

LEMMA 1. Let $\hat{u}$ be a first order efficient estimator with the second fundamental tensor $\hat{H}$. It holds that

$$(2.4)\quad \lim_{N\to\infty} E\, N[N\, e_i(\hat{u})e_j(\hat{u})g^{ij} - (n-m)]$$
$$= -\frac{1}{4}\|\overset{m}{\Gamma}\|^2 + \|\overset{e}{H}\|^2 + \frac{1}{4}\|\hat{H}\|^2 - 2(\overset{e}{H},\overset{m}{H}) - (\overset{m}{H},T),$$

where $\|\overset{m}{\Gamma}\|^2 \equiv \overset{m}{\Gamma}{}^a_{bc}\, \overset{m}{\Gamma}{}^d_{ef}\, \tilde{g}_{ad}\, \tilde{g}^{be}\, \tilde{g}^{cf}$,

$\|\overset{e}{H}\|^2 \equiv \overset{e}{H}{}^a_{b\lambda}\, \overset{e}{H}{}^c_{d\mu}\, \tilde{g}_{ac}\, \tilde{g}^{bd}\, \tilde{g}^{\lambda\mu}$,

$\|\hat{H}\|^2 \equiv \hat{H}^a_{\lambda\mu}\, \hat{H}^b_{\nu\xi}\, \tilde{g}_{ab}\, \tilde{g}^{\lambda\nu}\, \tilde{g}^{\mu\xi}$,

$(\overset{e}{H},\overset{m}{H}) \equiv \overset{e}{H}{}^a_{b\lambda}\, \overset{m}{H}{}^c_{d\mu}\, \tilde{g}_{ac}\, \tilde{g}^{bd}\, \tilde{g}^{\lambda\mu}$ and

$(\overset{m}{H},T) \equiv \overset{m}{H}{}^a_{b\lambda}\, T^c_{d\mu}\, \tilde{g}_{ac}\, \tilde{g}^{bd}\, \tilde{g}^{\lambda\mu}$,

with the tensor $T \equiv \overset{m}{\Gamma} - \overset{e}{\Gamma}$.

PROOF. The statistic $e_i(\hat{u})$ is expanded as

$$(2.5)\quad e_i(\hat{u}) = B_{\lambda i}\, e^\lambda + B_{ai}\Delta^a - \frac{1}{2}\partial_b B_{ai}\, e^a e^b - \partial_b B_{ai}\, e^a \Delta^b$$
$$- \frac{1}{6}\partial_c \partial_b B_{ai}\, e^a e^b e^c + O(\|e\|^4)$$

by Taylor's theorem, where $\Delta^a \equiv e^a - \bar{u}^a$. It follows from (2.2) that

$$(2.6)\quad \Delta^a = \frac{1}{2}\overset{m}{\Gamma}{}^a_{bc}\, e^b e^c - \overset{e}{H}{}^a_{b\lambda}\, e^b e^\lambda + \frac{1}{2}\hat{H}^a_{\lambda\mu}\, e^\lambda e^\mu + O(\|e\|^3).$$

Substituting (2.6) into (2.5), we have

$$(2.7)\quad e_i(u)e_j(u)g^{ij} = \tilde{g}_{\lambda\mu}\, e^\lambda e^\mu + \tilde{g}_{ab}\, \Delta^a\Delta^b + \frac{1}{4}\partial_b B_{ai}\, g^{ij}$$
$$\partial_d B_{cj}\, e^a e^b e^c e^d - \overset{m}{H}_{ab\lambda}\, e^a e^b e^\lambda + 2\overset{m}{H}_{ab\mu}\, \Delta^a e^b e^\mu$$
$$- \frac{1}{3} B^i_\lambda\, \partial_c \partial_b\, B_{ai}\, e^a e^b e^c e^\lambda - \overset{m}{\Gamma}_{abc}\, \Delta^a e^b e^c + O(\|e\|^5)$$
$$= \tilde{g}_{\lambda\mu}\, e^\lambda e^\mu + \frac{1}{2}\overset{m}{\Gamma}{}^a_{bc}\, \overset{m}{\Gamma}{}^d_{ef}\, \tilde{g}_{ad}\, e^b e^c e^e e^f + \overset{e}{H}{}^a_{c\lambda}\, \overset{e}{H}{}^b_{d\mu}$$
$$\times\, \tilde{g}_{ab}\, e^c e^d e^\lambda e^\mu + \frac{1}{4}\hat{H}^a_{\lambda\mu} H^b_{\nu\xi}\, e^\lambda e^\mu e^\nu e^\xi + \overset{m}{H}_{ab\mu}\, \overset{m}{H}_{cd\lambda}$$
$$\times\, \tilde{g}^{\lambda\mu}\, e^a e^b e^c e^d - \overset{m}{H}_{ab\lambda}\, e^a e^b e^\lambda + \overset{m}{H}_{ab\lambda}\, \overset{m}{\Gamma}{}^a_{cd}\, e^b e^c e^d e^\lambda$$

$$- 2\overset{m}{H}_{ab\mu}\ \overset{e}{H}{}^{a}_{c\lambda}\ e^b e^c e^\mu e^\lambda + \overset{m}{H}_{ab\lambda}\ \hat{H}^{a}_{\mu\xi}\ e^b e^\lambda e^\mu e^\xi +$$

$$- \frac{1}{3} B^i_\lambda \partial_c \partial_b B_{ai}\ e^a e^b e^c e^\lambda - \frac{1}{2} \overset{m}{\Gamma}_{abc}\ \overset{m}{\Gamma}{}^{a}_{df}\ e^b e^c e^d e^f -$$

$$+ \overset{m}{\Gamma}_{abc}\ \overset{e}{H}{}^{a}_{d\lambda}\ e^b e^c e^d e^\lambda + O(||e||^5)$$

since

$$g^{ij} = B^i_a\ \tilde{g}^{ab}\ B^j_b + B^i_\lambda\ \tilde{g}^{\lambda\mu}\ B^j_\mu\ .$$

Hence from (2.7) and Appendix II,

$$E[e_i(\hat{u})e_j(\tilde{u})g^{ij}] = \frac{n-m}{N} + \frac{1}{N^2}\{-\frac{1}{4}||\overset{m}{\Gamma}||^2 + ||\overset{e}{H}||^2 + \frac{1}{4}||\hat{H}||^2$$

$$- 2(\overset{e}{H},\overset{m}{H}) - (\overset{m}{H},T)\} + O(N^{-3}).$$

This completes the proof.

Now we set about

PROOF of THEOREM 2. Let $\tilde{u}$ be a first order efficient estimator. Then the statistic $\rho_{KL}(\hat{x}, \eta(\tilde{u}))$ is expanded as

(2.8) $$\rho_{KL}(\hat{x}, \eta(\tilde{u})) = \frac{1}{2}\tilde{e}_i\tilde{e}_j g^{ij} - \frac{1}{2} T^{ijk}\tilde{e}_i\tilde{e}_j e_k + \frac{1}{3} T^{ijk}\tilde{e}_i\tilde{e}_j\tilde{e}_k$$

$$+ \frac{3}{4} S^{ijk\ell}\tilde{e}_i\tilde{e}_j e_k e_\ell + S^{ijk\ell}\tilde{e}_i\tilde{e}_j\tilde{e}_k e_\ell + \frac{1}{8} S^{ijk\ell}\tilde{e}_i\tilde{e}_j\tilde{e}_k\tilde{e}_\ell$$

$$+ O(||e||^5),$$

where $S^{ijk\ell} \equiv \frac{\partial}{\partial\eta_i} T^{jk\ell}$ and $\tilde{e}_i \equiv e_i(\tilde{u})$. By a similar argument as in the proof of LEMMA 1, we have from (2.8) that

(2.9) $$\rho_{KL}(\tilde{x}, \eta(\tilde{u})) = \frac{1}{2}\tilde{e}_i\tilde{e}_j g^{ij} - \frac{1}{2} T_{\lambda\mu i}\ e^\lambda e^\mu e^i$$

$$- T_{a\lambda i}\ \Delta^a e^\lambda e^i + \frac{1}{3} T_{\lambda\mu\nu}\ e^\lambda e^\mu e^\nu + T_{\lambda\mu a}\ e^\lambda e^\mu \Delta^a$$

$$- \frac{3}{4} S_{\lambda\mu ij}\ e^\lambda e^\mu e^i e^j + S_{\lambda\mu\nu i}\ e^\lambda e^\mu e^\nu e^i + \frac{1}{8} S_{\lambda\mu\nu\xi}\ e^\lambda e^\mu e^\nu e^\xi$$

$$+ O(||e||^5)$$

Let $\tilde{H}$ be the second fundamental tensor of the ancillary subspace of $\tilde{u}$. It follows from (2.9) and LEMMA 1 that

$$(2.10)\quad E\,\rho_{KL}(\hat{\theta},\ \theta(\tilde{u})) = \frac{n-m}{N} + \frac{1}{N^2}\{\tfrac{3}{8}||\tilde{H}||^2 + M\} + O(\frac{1}{N^3}),$$

where

$$M = -\tfrac{1}{8}||\overset{m}{\Gamma}||^2 + \tfrac{1}{2}||\overset{e}{H}||^2 - \tfrac{1}{2}(\overset{m}{H},T) - (\overset{e}{H},\overset{m}{H})$$
$$-\tfrac{1}{6}\,T_{\lambda\mu\nu}\,T^{\lambda\mu\nu} - \tfrac{1}{2}\,T_{\lambda\mu a}\,T^{\lambda\mu a}$$
$$+\tfrac{15}{8}\,S_{\lambda\mu\nu\xi}\,\tilde{g}^{\lambda\mu}\,\tilde{g}^{\nu\xi} + \tfrac{3}{4}\,S_{\lambda\mu ab}\,\tilde{g}^{ab}\,\tilde{g}^{\lambda\mu}.$$

Clearly the term M in the RHS of (2.10) is independent of $\tilde{u}$ and dependent only on the model $\tilde{\mathfrak{F}}$. Hence it holds for a second order efficient estimator $\hat{u}$ that

$$(2.11)\quad E[\rho_{KL}(\hat{\theta},\ \theta(\tilde{u}) - \rho(\hat{\theta},\ \theta(\hat{u}))] = \frac{1}{N^2}\,\frac{3}{8}||\tilde{H}||^2 + O(\frac{1}{N^3})$$

since the estimator $\hat{u}$ has the vanishing second fundamental tensor on account of THEOREM A. Therefore the second order efficiency of $\hat{u}$ implies the inequality (1.2). The inverse assertion is clear since $||\hat{H}|| = 0$ implies $\hat{H} = 0$. This completes the proof.

Similary we have

PROOF of THEOREM 3. The statistic $\rho(\hat{\theta},\ \theta(\tilde{u}))$ is expanded as

$$(2.12)\quad \tfrac{1}{2}\,g^{(\rho)\ ij}(\hat{\theta})\tilde{e}_i\tilde{e}_j + \tfrac{1}{6}\{2\,{}^{*}\Gamma^{(\rho)\ ij|k}(\hat{\theta}) + \Gamma^{(\rho)ij|k}(\hat{\theta})\}\tilde{e}_i\tilde{e}_j\tilde{e}_k$$
$$+ D^{(\rho)ijk\ell}\tilde{e}_i\tilde{e}_j\tilde{e}_k\tilde{e}_\ell + O(||e||^5),$$

where $A(\rho) \equiv (g^{(\rho)},\ \Gamma^{(\rho)},\ {}^{*}\Gamma^{(\rho)})$ and

$$D^{(\rho)ijk\ell} \equiv \frac{\partial^4}{\partial\eta_i\partial\eta_j\partial\eta_k\partial\eta_\ell}\,\rho(\theta[\eta],\ \theta(u))|_{\eta=\eta(u)}.$$

If $A(\rho) = A$ on $\tilde{\mathfrak{F}}$, the expansion (2.12) is equal to (2.8) on ρ_{KL} except the last term because of $A(\rho_{KL}) = A$. This implies the condition (2.11) with ρ_{KL} replaced by ρ, which completes the proof.

REMARK 1. From the first term of (2.12), it follows that Theorem 1 holds for ρ with $g^{(\rho)} = g$ on $\tilde{\mathfrak{F}}$ in place of ρ_{KL}.

Let ρ be a contrast function on $\mathfrak{F}$ with the almost-metric structure $A(\rho)$. By a similar argument as in the proof of THEOREMS 2 and 3, we may conclude the relation

$$\lim_{N\to\infty} N^2\, E[\rho(\hat{\theta}, \tilde{u}) - \min_{u\in U} \rho(\hat{\theta}, \theta(u))] = \frac{3}{8}||H^{(\rho)} - \tilde{H}||^2$$

for any first order efficient estimator $\tilde{u}$, where $\tilde{H}$ and $H^{(\rho)}$ denote the second fundamental tensors of the ancillary subspace of $\tilde{u}$ and the subspace

$$\{\, f(\cdot|\theta) \;;\; \rho(\theta,\theta(u)) = \min_{u'\in U} \rho(\theta,\theta(u'))\,\}$$

at $f = f(\cdot|\theta(u))$, respectively, with respect to $\overset{m}{\Gamma}$. Finally

$$\lim_{N\to\infty} N^2\, E[(\tilde{u}^a - u_\rho^a)\tilde{g}_{ab}(\tilde{u}^b - u_\rho^b)]$$

$$= \lim_{N\ \infty} N^2\, E[\rho(\hat{\theta}, \theta(\tilde{u})) - \min_{u\in U} \rho(\hat{\theta}, \theta(u))],$$

where $\hat{u}_\rho$ denotes the minimum contrast estimator based on ρ.

Acknowledgements. I would express my hearty thanks to Professor M. Okamoto and Dr. Y. Toyooka of Osaka University.

Appendix I

Let τ be a co-ordinate system (parameter vector) of $\mathfrak{F}$ with the transformation $\theta\{\cdot\}$ of τ into θ. The log-likelihood $\log f(x|\theta\{\tau\})$ with respect to τ is denoted by $\ell(\tau)$. Then the almost-metric structure $A = (g, \overset{m}{\Gamma}, \overset{e}{\Gamma})$, introduced by Amari [1], is defined as the following components

$$g_{ij}(\tau) = E[\frac{\partial\ell}{\partial\tau^i}\,\frac{\partial\ell}{\partial\tau^j}]\ ,$$

$$\overset{m}{\Gamma}_{ij|k}(\tau) = E[\frac{\partial^2\ell}{\partial\tau^i\partial\tau^j}\,\frac{\partial\ell}{\partial\tau^k}] + E[\frac{\partial\ell}{\partial\tau^i}\,\frac{\partial\ell}{\partial\tau^j}\,\frac{\partial\ell}{\partial\tau^k}]$$

and

$$\overset{e}{\Gamma}_{ij|k}(\tau) = E[\frac{\partial^2 \ell}{\partial\tau^i \partial\tau^j} \frac{\partial \ell}{\partial\tau^k}],$$

with respect to τ with $\ell = \ell(\tau)$ (see also Dawid [2]).

On the other hand, the almost-metric structure associated with a contrast function ρ on $\mathfrak{F}$ is defined as the following components

$$g_{ij}^{(\rho)}(\tau) = -\frac{\partial^2}{\partial\tau_1^i \partial\tau_2^j} \rho(\tau_1, \tau_2)|_{\tau_1=\tau_2=\tau} ,$$

$$\Gamma_{ij|k}^{(\rho)}(\tau) = -\frac{\partial^3}{\partial\tau_1^i \partial\tau_1^j \partial\tau_2^k} \rho(\tau_1, \tau_2)|_{\tau_1=\tau_2=\tau}$$

and

$${}^*\Gamma_{ij|k}^{(\rho)}(\tau) = -\frac{\partial}{\partial\tau_1^i \partial\tau_1^j \partial\tau_2^k} \rho(\tau_2, \tau_1)|_{\tau_2=\tau_1=\tau} ,$$

with $\rho(\tau_1, \tau_2) = \rho(f(\cdot|\theta\{\tau_1\}), f(\cdot|\theta\{\tau_2\}))$ (c.f. Eguchi [4]).

Appendix II

It holds for any sample size N that

$$E\, e^a e^b = \frac{1}{N} \tilde{g}^{ab} ,$$

$$E\, e^\lambda e^\mu = \frac{1}{N} \tilde{g}^{\lambda\mu} ,$$

$$E\, e^a e^b e^\lambda = \frac{1}{N^2} T^{ab\lambda} ,$$

$$E\, e^a e^b e^\lambda e^\mu = \frac{1}{N^2} \tilde{g}^{ab} \tilde{g}^{\lambda\mu} + \frac{1}{N^3} S^{ab\lambda\mu}$$

and

$$E\, e^a e^b e^c e^\lambda = \frac{1}{N^3} S^{abc\lambda} ,$$

where

$$S^{ab\lambda\mu} \equiv B^{ai} B^{bj} B^{\lambda k} B^{\mu\ell}\, \partial_i \partial_j \partial_k \partial_\ell \psi$$

with $\partial_i \equiv \partial/\partial\theta^i$.

References

[1] Amari,S. (1982). Differential geometry of curved exponential families - curvature and information loss. Ann. Statist. 10, 357-385.

[2] Dawid,A.P. (1975). Comments on the paper by Efron. Ann. Statist. 3, 1231-1234.

[3] Efron,B. (1975). Defining the curvature of a statistical problem. Ann. Statist. 3, 1189-1217.

[4] Eguchi,S. (1983). Second order efficiency of minimum contrast estimators in a curved exponential family. (Forthcoming)

[5] Eguchi,S. (1983). Contrast functions of Mahalanobis-type on an exponential family. (Submitted).

[6] Kuboki,H. (1982). Unbiased estimations in the sense of Lehmann and their discrimination rates. Ann. Inst. Statist. Math. 34, A, 19-37.

[7] Ghosh,J.K. and Subramanyan,K. (1974). Second order efficiency of maximum likelihood estimators. Sankyā A, 36, 325-358.

[8] Rao,C.R. (1963). Criteria of estimation in large samples. Sankyā A, 25, 189-206.

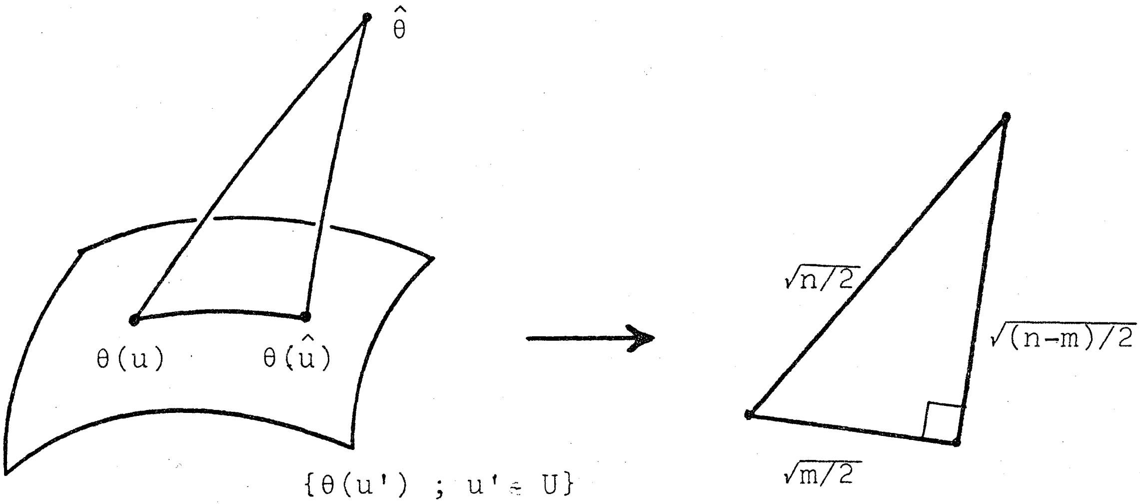

Fig. 1. The right triangle. Let $\hat{u}$ be first order efficient estimator of u. The triange with sides $\sqrt{N\ E\ \rho(\hat{\theta},\theta(u))}$, $\sqrt{N\ E\ \rho(\hat{\theta},\theta(\hat{u}))}$ and $\sqrt{N\ E\ \rho(\theta(\hat{u}),\theta(u))}$ converges to the right triangle with $\sqrt{n/2}$, $\sqrt{(n-m)/2}$ and $\sqrt{m/n}$.